

RFC 7990 : RFC Format Framework

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 16 décembre 2016

Date de publication du RFC : Décembre 2016

<https://www.bortzmeyer.org/7990.html>

Voici enfin la série de RFC décrivant le nouveau format des RFC. Ce projet a commencé il y a plusieurs années, mais les discussions ont été longues. Ce nouveau RFC et les huit autres qui l'accompagnent, marquent un changement important dans ces « textes sacrés » de l'Internet : l'ancien format « texte brut » n'est plus le format de référence. Désormais, tout RFC sera fait en XML, format d'où seront produits automatiquement des versions texte brut, HTML, PDF, etc.

Les RFC sont des documents cruciaux pour l'Internet. C'est sous forme de RFC que sont publiées les normes techniques de la famille de protocoles TCP/IP. Et il y a bien d'autres RFC qui ne sont pas forcément des normes (cf. RFC 1796¹). Librement disponibles en ligne (contrairement aux normes techniques des organisations traditionnelles du mésozoïque), dans un format ouvert, la disponibilité des RFC est l'une des raisons du succès de l'Internet. (Ce RFC sur le format a depuis été remplacé par le RFC 9720.)

Parlons de format, justement. Les RFC, jusqu'à maintenant, étaient sous forme de texte brut, et en ASCII seul. (Certes, des versions PDF et HTML non-officielles étaient diffusées mais on voyait bien qu'elles avaient été produites à partir du texte brut... Certes, il était possible depuis dix ans d'écrire les RFC en XML, cf. RFC 2629, mais ce n'était pas le format de référence.) Pourquoi donc <<https://www.bortzmeyer.org/rfc-en-texte-brut.html>> se limiter à ce format ? Il y avait plusieurs bonnes (et d'autres moins bonnes) raisons mais il ne faut pas de cacher qu'une des raisons les plus importantes était qu'il est difficile de faire changer un processus de production bien établi, même s'il comprend des archaïsmes, comme l'utilisation de troff pour traiter des documents. L'actuelle éditrice des RFC, Heather Flanagan, a donc eu bien du mérite à faire aboutir ce projet de changement. Il a fallu beaucoup de discussions, dans une communauté souvent divisée. (Que les informaticiens pensent aux grands débats du genre « vi ou emacs ? »)

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc1796.txt>

Le projet de réforme des RFC avait sérieusement commencé en 2013 avec le RFC 6949, le véritable cahier des charges du nouveau format. La décision formelle de migrer vers le nouveau format, et donc de décider que le format de référence serait désormais le XML et non plus le texte brut a été prise en mai 2013 <<http://www.rfc-editor.org/pipermail/rfc-interest/2013-May/005584.html>>. Avant et pendant cette décision, d'innombrables messages ont été échangés sur la liste de diffusion rfc-interest <<https://www.rfc-editor.org/mailman/listinfo/rfc-interest>>.

Il est important de noter que cette discussion portait sur le processus de publication des RFC terminés. L'élaboration des "*Internet-Drafts*", la décision de les publier ou pas (qui dépend de chaque voie, cf. RFC 8729) ne sont pas concernées.

La section 2 de notre RFC résume le problème que veut résoudre le nouveau format. Le monde a bien changé depuis que seuls une poignée de Californiens anglophones venait aux réunions IETF. Les participants viennent aujourd'hui de 45 pays, situés dans le monde entier, les lecteurs des RFC sont plus divers que jamais, utilisent des engins très variés, et il est toujours aussi crucial que les RFC soient largement disponibles et accessibles, et sur une très longue période (mon favori est le RFC 768, publié en 1980 et toujours d'actualité). Le format de référence en texte ASCII brut ne permettait clairement pas cela.

Mais choisir un successeur n'était pas facile : notre RFC insiste sur le fait qu'il y a aujourd'hui plusieurs groupes qui utilisent les RFC (pas seulement des techniciens, juristes et chefs accèdent aujourd'hui à des RFC), et sur le fait qu'il fallait un compromis entre les besoins actuels et l'importance d'une disponibilité à long terme (par exemple, adopter le format à la mode du moment pourrait se payer cher plus tard lorsque ce format n'intéressera plus personne).

Un peu de terminologie (section 3) est nécessaire pour bien comprendre le choix effectué :

- Format canonique : le format de référence, archivé, utilisé en cas de conflit (XML, donc). Il est important d'en avoir un, au cas où une bogue dans les logiciels change une partie d'un RFC.
- Formats de publication : les divers formats sous lesquels est publié le RFC (détaillés en section 7). Peu de gens liront le RFC en XML (quoique cela soit possible, une propriété qui est importante pour la conservation à long terme). Ils liront de l'HTML, du PDF, voire du texte seul pour les traditionalistes. Tous ces formats seront produits (de préférence automatiquement) à partir du format canonique. Chacun a ses avantages et inconvénients (section 5). Par exemple, HTML avec JavaScript fournit des capacités de navigation bien meilleures que le texte brut.
- Format révisable : le format que le "*RFC Editor*" utilisera pour son travail interne. Ce sera XML.
- Format de soumission : le format sous lequel le texte sera transmis initialement par les auteurs au "*RFC Editor*". Aujourd'hui, le texte brut est obligatoire, le XML autorisé en prime. Demain, ce sera du XML, mais avec des exigences moins strictes que pour le format canonique.

Et aussi un terme important : texte réagencable (ou réajustable, "*reflowable text*"). C'est du texte qui s'ajuste automatiquement à la largeur du dispositif de lecture. C'est banal dans le monde HTML, où c'est fait automatiquement depuis toujours. Mais c'était un des principaux inconvénients de l'ancien format des RFC : le texte avait une largeur fixe.

Quel sera donc exactement le format canonique? La section 6 répond à cette question :

- Le langage sera du XML avec le vocabulaire spécifié dans le RFC 7991 (nommé « v3 »). Il est normalement meilleur que les précédents vocabulaires utilisés depuis le RFC 2629.
- Les auteurs pourront envoyer leur "*draft*" en suivant le format dit « v2 » (celui du RFC 7749), voire en texte brut, mais il sera ensuite converti dans le format v3 ci-dessus.
- Le SVG sera autorisé dans le source XML.
- Comme en v2, DTD est abandonné, la description officielle du schéma XML est en Relax NG.
- Les textes obligatoires (RFC 5741) seront automatiquement insérés.

- Le source XML du format canonique sera autonome. Cela veut dire qu'il n'aura pas de références à des sources extérieures. Ainsi, si un auteur référence un code source externe avec `<sourcecode src="[un URI externe]" . . .` (RFC 7991, section 2.48), le code en question sera inclus dans la version canonique. Idem si l'auteur a utilisé XInclude.
- Il n'y aura pas de commentaires ou de "*processing instructions*" dans le source XML. Si l'auteur en a mis, ils seront retirés.

Notez donc que les images (en SVG) seront désormais possibles (voir le RFC 7996).

Le guide du style des RFC (RFC 7322) avait été révisé pour tenir compte de ce nouveau format. Notamment, il se concentre désormais sur le contenu du texte, ne demandant plus aux auteurs des efforts de présentation. (La section 5 résume les changements importants pour les auteurs.)

Enfin, la section 7 décrit les formats de publication. À partir du source XML, seront automatiquement produits HTML, PDF, texte brut et peut-être plus tard d'autres. Le HTML est évidemment la cible évidente. Son utilisation pour les RFC est décrite dans le RFC 7992. Le résultat sera certainement bien meilleur que les versions HTML non-officielles actuelles, qui sont produites à partir du texte brut, qui ne contient pas assez de structure pour faire du bon HTML. La mise en page sera évidemment assurée par CSS (RFC 7993), il y aura une feuille de style standard, que chacun sera bien sûr libre de remplacer. Le SVG sera inclus dans l'HTML (il faudra donc un navigateur qui gère bien SVG). Il y aura même du JavaScript mais avec de sévères restrictions. Notamment, le code JavaScript ne devra pas changer le texte, ou supprimer du texte.

PDF, quant à lui, est spécifié dans le RFC 7995. Il devra suivre le profil PDF/A-3, spécialement prévu pour de l'archivage à long terme, et pour pouvoir être relu par des logiciels PDF n'ayant pas tous les derniers gadgets.

Naturellement, le texte brut n'est pas abandonné. Comme indiqué dans le RFC 7994, il y aura une version en texte brut produite automatiquement à partir du XML, même si elle ne sera plus la version canonique. Parmi les nouveautés par rapport à l'ancien format, UTF-8 sera désormais autorisé, même si c'est de façon limitée (voir les limitations dans le RFC 7997). Il pourra y avoir une variante non découpée en pages.

Dans le futur, il est possible que le format EPUB soit ajouté à cette liste.

Au passage, comment a été décidé cet important changement dans le format des RFC ? La section 4 résume cette histoire. Comme indiqué plus haut, cela a pris très longtemps et nécessité beaucoup de discussions, qui ont notamment eu lieu sur la liste de diffusion rfc-interest `<https://www.rfc-editor.org/mailman/listinfo/rfc-interest>`, et au cours des réunions physiques de l'IETF. Le cahier des charges a été formalisé en 2013 dans le RFC 6949. Une fois le cahier des charges décidé, une équipe spécialisée a été désignée par le "*RFC Editor*" pour mettre au point les détails, notamment en adaptant le langage XML utilisé, partant de la dernière version (RFC 7749), pour arriver au futur langage, RFC 7991. Des éditeurs professionnels ont également été consultés, ainsi d'autres SDO et même des juristes (oui, car aux États-Unis, rien n'est désormais à l'abri d'actions en justice, même pas les RFC, le choix du format de sortie PDF/A-3 venait en partie de la nécessité de répondre aux "*subpoenas*"). Le tout était bien sûr fait sous la supervision du "*RFC Series Oversight Committee*" `<https://www.iab.org/activities/programs/rfc-editor-program/>`. Certaines décisions furent consensuelles, les autres tranchées par le "*RFC Editor*" (cf. RFC 8728). Le tout a été approuvé par l'IAB en août 2016 `<https://www.iab.org/2016/08/06/iab-approves-rfc-format-related-drafts-for-publication/>`.

Après ce tour du passé, le futur. Comment se fera la transition vers le nouveau système (section 10) ? C'est qu'il va falloir créer de nouveaux outils (cf. RFC 7998). L'appel d'offres pour leur développement a

été fait en septembre 2016 <<https://mailarchive.ietf.org/arch/msg/ietf-announce/pC618gbTXrLIIA>>
La description des outils est une très intéressante lecture <<http://www.nostrum.com/~rjspark/rfced/>> (l'appel d'offres formel est sur la page des Request For Proposal <<https://iaoc.ietf.org/rfps.html>>). L'appel d'offres a été gagné par les sociétés SeanTek et Elf Tools.

Pendant une période intermédiaire, le texte seul sera toujours utilisé comme format canonique, mais les nouveaux RFC passeront également par le nouveau "*workflow*", pour vérifier que tout se passe bien et que le résultat est correct. Double travail, donc, mais nécessaire pour s'assurer que tout est en place.

Notez que, même une fois la transition finie, les auteurs ne seront pas **forcés** de soumettre leur document sous forme d'un fichier XML (ils seront simplement très fortement encouragés à le faire). S'ils envoient le texte seul comme avant, le "*RFC Editor*" devra produire le XML lui-même, et c'est ce XML qui sera la version canonique. Rappelez-vous que beaucoup de RFC sont des documents normatifs et que chaque mot, voire chaque virgule peut compter! Voici pourquoi il faudra s'assurer que tout est parfait, même si, au début, cela entraînera certainement des retards dans la publication.

Dans le cas où l'auteur envoie du XML suivant le RFC 7991, il y aura moins de travail pour le "*RFC Editor*", juste convertir ce XML au XML canonique (résoudre les références extérieures, par exemple) et passer ce XML canonique dans les nouveaux outils.

Notez que le "*RFC Editor*" maintient une FAQ très utile <<https://www.rfc-editor.org/rse/format-faq/>> sur toutes les questions que pose le nouveau format. Et la "*RFC Editor*" avait fait un très drôle Pecha Kucha à Séoul en novembre 2016, sur le cahier des charges du nouveau format <<http://snaggletooth.akam.ai/Seoul-videos/7-heather-8.mp4>>.

Le premier RFC au nouveau format a été le RFC 8651, sorti en octobre 2019.